

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 02:53:24 ; Search time 327 Seconds
(without alignments)
4977.863 Million cell updates/sec

Title: US-09-744-315-7

Sequence: 603
1 caggaacagcctctcctgc.....aaaataaagtcctgc 603

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq.19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	603	100.0	603	AAZ50577	Human epidermal pr
2	511.4	84.8	666	AAK51648	Human polynucleoti
3	509.8	84.5	617	AAK52632	Human polynucleoti
4	355.4	58.9	486	21 AACT7838	Human cancer assoc
5	142.6	23.6	689	24 ABT09264	Phase-1 Rat CT gen
6	134.4	22.3	148	24 ABRK5393	Human head and nec
7	104.8	17.4	422	22 AAK52162	Human polynucleoti
8	104.8	17.4	422	22 AAK53146	Human polynucleoti

9	95.6	15.9	540	23 AAK56924	DNA encoding novel
10	93.8	15.6	414	22 AAK52164	Human polynucleoti
11	93.8	15.6	414	22 AAK53148	Human polynucleoti
12	85.8	14.2	695	22 AAK81731	Human secreted pro
13	72.4	12.0	566	22 AAK85555	CDNA encoding CDF
14	71	11.8	528	21 AAZ50579	Human epidermal pr
15	68.8	11.4	678	24 ABO61174	Human epidermal pr
16	62.4	10.3	697	21 AAZ50578	Human epidermal pr
17	44.6	7.4	403	21 ABR10240	Trinucleotide repe
18	43.4	7.2	4335	23 ABR19854	DNA encoding novel
19	43.2	7.2	397	20 AAK89891	Spinozerbellar at
20	43.2	7.2	1159	21 AAK59240	An Ecorti fragment
21	42.6	7.1	844	24 ABR95177	Gene #2215 used to
22	42.4	7.0	515	20 AAX18930	Human huntingtin g
23	42.4	7.0	600	25 ABE28178	Huntington's disea
24	42.4	7.0	614	25 ABE281762	Huntington's disea
25	42.4	7.0	10348	20 AAZ23426	Human huntingtin p
26	42.4	7.0	10348	22 ABR06651	Human huntingtin p
27	42.4	7.0	10348	23 AAS91971	DNA encoding novel
28	42.4	7.0	10366	15 AAK67401	Huntingtin DVA\ITI
29	42.4	7.0	10366	18 AAT97924	Composite sequence
30	42.4	7.0	10366	19 AAV05828	Human huntingtin c
31	42.2	7.0	477	21 AAZ44307	Human SCAT genomic
32	42	7.0	1551	24 ABR19450	Mouse ischaemic co
33	42	7.0	1551	25 ABR76651	Gene #1748 used to
34	41.8	6.9	2614	24 ABRK36100	CDNA sequence #491
35	41.8	6.9	6604	24 ABR95250	Human CDNA difere
36	41.8	6.9	6788	24 ABRK3506	Antifreeze peptide
37	41.6	6.9	6794	21 AAZ45597	Synthetic antifree
38	41.6	6.9	336	20 AAZ00010	Human zsl963 degen
39	41.6	6.9	336	20 AAX8092	Human secreted sal
40	41.6	6.9	657	24 AAD45051	Human zsl963 degen
41	41.6	6.9	657	24 ABR52634	Human zsl963 degen
42	41.6	6.9	657	24 AAS20592	Human zsl963 degen
43	41.6	6.9	657	25 ABR93595	Human zsl963 degen
44	41.6	6.9	1037	21 AAK59241	Exons E, C and A o
45	41.6	6.9	1472	21 AAK59241	Exons D, C, B and

ALIGNMENTS

RESULT 1	
AAZ50577	AAZ50577 standard; cDNA; 603 BP.
ID	AAZ50577
AC	AAZ50577;
XX	23-MAY-2000 (first entry)
XX	Human epidermal protein-1 cDNA.
XX	Human epidermal protein-1; HEP1; epithelial disorder; scabies;
XX	dyshidrotic eczema; cell proliferative disorder; actinic keratosis;
XX	arteriosclerosis; autoimmune disorder; inflammatory disorder;
XX	acquired immune deficiency syndrome; AIDS; Addison's disease; antiHIV;
XX	dematological; antiteriosclerotic; antiinflammatory;
XX	immunosuppressive; ss.
XX	Homo sapiens.
XX	OS
XX	Key
XX	Location/Qualifiers
XX	56..334
XX	/*tag= a
XX	/product= "Human epidermal protein-1"
XX	CDs
XX	WO200006727-A2.
XX	10-FEB-2000.
XX	27-JUL-1999;
XX	99MO-US17107.
XX	28-JUL-1998;
XX	98US-0155203.

Qy	1	CAGGAACAGCTTCTCTGCTCTCTTGACCTGGACCAATCAATCTCTGC	CAAAGTGC	60
Db	43	CAGGAACAGCTTCTCTGCTCTCTTGACCTGGACCAATCAATCTCTGC	CAAAGTGC	102
Qy	61	CTGCGCAGCAACCAAGAGCAATGCCAACCCCAACCAAGTCTCCCTCA	CCCAAGTCTC	120
Db	103	CTGCGCAGCAACCAAGAGCAATGCCAACCCCAACCAAGTCTCCCTCA	CCCAAGTCTC	167
Qy	121	CCCAAGAGCCAGTACAGTGTGTGCTCAGCTTCTCTGACTGTGCCCCA	AGCTCTGG	180
Db	163	CCCAAGAGCCAGTACAGTGTGTGCTCAGCTTCTCTGACTGTGCCCCA	AGCTCTGG	222
Qy	181	GGTCTGTGGCCCTAAGCTCCGAGGGGGGGCTTCTTGAAACAACAAG	GGCCACACCG	240
Db	223	GGGCTGTGGCCCTAAGCTCCGAGGGGGGGCTTCTTGAAACAACAAG	GGCCACACCG	287
Qy	241	ATGCGCGCGCCAGAGGCCCAATCTCTGTACAAGGGCAATGTGTACGA	AGCGGGGGCTC	300
Db	283	ATGCGCGCGCCAGAGGCCCAATCTCTGTACAAGGGCAATGTGTACGA	AGCGGGGGCTC	342
Qy	301	TGGGTGTGTGCAAGGTCTTGCGGGGGCGTGGCTGAATCAGATCTGA	TGTGACAAGCG	360
Db	343	TGGGTGTGTGCAAGGTCTTGCGGGGGCGTGGCTGAATCAGATCTGA	TGTGACAAGCG	407
Qy	361	ATCTTTGAGAGAAACAAGATTCCAAGAGGCCAAGAACGCCCATCTGA	CGCATGCTT	420
Db	403	AT-TTTGAGAGAAACAAGATTCCAAGAGGCCAAGAACGCCCATCTGA	CGCATGATG	461
Qy	421	CCCATATACCTCTCTTGACTTCAAGAGGTAGCTGAGAGTTTCTCTGT	GGGGATCTG	480
Db	462	CCCATATACCTCTCTCTTGACTTCAAGAGGTAGCTGAGAGTTTCTCT	GTGGAAGTGTG	522
Qy	481	AGCTCTCCCAAGAGGCACTTCTGTTTATGTACAGATGTCAATGTCCC	CTTACCCC	540
Db	522	AGCTCTCCCAAGAGGCACTTCTGTTTATGTACAGATGTCAATGTCCC	CTTACCCC	581
Qy	541	TGTACTCGCAAGATTTGGCAGTGTGTGGCCAACTGTAAAGATTAAG	TTTC	597
Db	582	TGTACTCGTAAAGATTGGCAGTGTGTGGCCAACTGTAAAGATTAAG	TTTC	638
RESULT 3				
AAKS2632				
ID	AAKS2632 standard; cDNA; 617 BP.			
AC	AAKS2632;			
XX				
DT	06-NOV-2001 (first entry)			
XX				
DE	Human polynucleotide SEQ ID NO 2161.			
XX				
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;			
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;			
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;			
KW	nervous system disorder; arthritis; inflammation; ss.			
XX				
OS	Homo sapiens.			
XX				
PN	WO200157190-A2.			
XX				
PD	09-AUG-2001.			
XX				
PF	05-FEB-2001; 2001MO-US04098.			
XX				
PR	03-FEB-2000; 2000US-0496914.			
PR	27-APR-2000; 2000US-0560875.			
PR	20-JUN-2000; 2000US-0598075.			
PR	19-JUL-2000; 2000US-0620325.			
PR	01-SEP-2000; 2000US-0654936.			
PR	15-SEP-2000; 2000US-0663361.			
PR	20-OCT-2000; 2000US-0683361.			
PR	30-NOV-2000; 2000US-0728422.			

PA	(HYSE-) HYSEQ INC.
XX	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
XX	Pi Zhao CA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI	Xue MJ, Yang T, Wejhrman T, Goodrich R,
XX	WPI; 2001-476283/51.
DR	F-PDB; AAM79499.
XX	Nucleic acids encoding polypeptides with cytokine-like activities,
PT	useful in diagnosis and gene therapy -
XX	Claim 1, Page 4528-4529; 622IP; English.
PS	The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC	encoded polypeptides (AAM78322-AAH80302) that exhibit activity elating to
CC	Cytokine, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cytokine-like activities,
CC	e.g., stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	Inflammation.
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC	(AAM80020) are omitted as the relevant pages from the sequence listing
XX	were missing at the time of publication.
SQ	Sequence 617 BP; 134 A; 198 C; 157 G; 128 T; 0 other;
Query Match	84.5%; Score 509.8; DB 22; Length 617;
Best Local Similarity	92.0%; Pred. No. 1,2e-127;
Matches 549; Conservative	0; Mismatches 47; Indels 1; Gaps 1;
OY	1 CAGGAACACCCTTCTCCTCGGTCCTGTGGCACTGGACAATCATCTGCAGACATGCC 60
Db	1 CAGGAACACCCTTCTCCTCGGTCCTGTGGCACTGGACAATCATCTGCAGACATGCC 60
OY	61 CTGCCAGAGAACAACAGCAGAGTGCCAACCCCACCCCAAAGTGTCTCTCACCCAAGTGC 120
Db	61 CTGCCAGAGAACAACAGCAGAGTGCCAACCCCACCCCAAAGTGTCTCTCACCCAAGTGC 120
OY	121 CGCAAGAACCCAGATAAGTAGTGTCTGTGACTCCAGACTTCTGTGGGTGGCCCCAAGCTGG 180
Db	121 CGCAAGAACCCAGATAAGTAGTGTCTGTGACTCCAGACTTCTGTGGGTGGCCCCAAGCTGG 180
OY	181 GGATTCTGGACCTTAAGTCCAGAGGCGGGTGTCTTTCTGAACACCAAGGCGCCACACCG 240
Db	181 GGAGCTGTGGCTTAAGTCCAGAGGCGGGTGTCTTTCTGAACACCAAGGCGCCACACCG 240
OY	241 ATGCGGGGCGCAGAGAGCCCACTCTGTGACAGGGGAGTGTCAACAAGGCGGGGCGTCC 300
Db	241 ATGCGGGGCGCAGAGAGCCCACTCTGTGACAGGGGAGTGTCAACAAGGCGGGGCGTCC 300
OY	301 TGCGTCTGTCCACGGTTCTGGGCGGTGCTGTATCCAGTCTGTATGCTGAGCAAGCG 360
Db	301 TGCGTCTGTCCATGCGTCTGAGAGCGTGTGTGTGATCTGTATGCTGAGCAAGCG 360
OY	361 ATCTTTGAGAGAAACAAGATCTCCAAAGAGGCAAGAACAGCCCATCTGACCATGCGTT 420
Db	361 AT-TTTGAGAGAAACAAGATCTCCAAAGAGGCCAGAAAACCCCATTTGATCDAGAGTT 419
OY	421 CCATATACCTCTTCTGTGACTTTCACAGGCTAGCTGAGAGTTTTTCTGTGGGAGATCTG 480
Db	420 CCCAGATACCTCTTCTGTGACTTTCACAGGCTAGCATNAGGGTTTTTCTGTGAGAGGTTCTG 479
OY	481 AGCTCTCCCAAGAGCACTTCTTTATGTATGACAGATGTCAATATGTCTCCCTACCC 540
Db	480 AGCTCTCCCAAGAGCACTTCTCTGTGATGTCAAGATGTCAATATGTCTCCCTTACCC 539
OY	541 TGTACTCGCAAGATTTGCAGTCTGTGTGCCCACTCGTAAAAAAGATTAAGTTCC 597

XX Homo sapiens.
 OS
 XX W0200061627-A1.
 PN
 XX
 PD 19-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US09067.
 XX
 PR 03-APR-1999; 99US-0126697.
 PR 20-JUN-2000; 2000US-0176829.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN G A.
 PI Rosen CA, Ruben SM, Komatsu G;
 XX
 DR WPI; 2000-647419/62.
 DR P-PSDB; AAB45407.
 XX
 XX New nucleic acid molecules encoding 49 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 1; Page 438; 521pp; English.

CC The invention relates to the isolation of genes A081710-C81758 encoding
 CC 49 human secreted proteins AAB45386-B54534. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (S0101) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 XX
 SQ Sequence 695 BP; 137 A; 230 C; 168 G; 157 T; 3 other;

Query Match 14.2%; Score 85.8; DB 21; Length 695;
 Best Local Similarity 59.5%; Pred. No. 3.2e-13;
 Matches 234; Conservative 1; Mismatches 98; Indels 60; Gaps 3;

QY 1 CAGGACAGCCTTCTCTGCTCTCTGACCTGACACTGACACTCTGCTGCAAGATGTC 60
 DB 66 CAGGACAGTGTGTGTGTCTCTGCTGTGACAGGGTGTACTAACTCTGCGCAGATGTC 125
 QY 61 CTGCGACCAAGACAG 107
 DB 126 TTGCGACCAAGACAG 185
 QY 108 -----CACCAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 159
 DB 186 AATATGTCACTAAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 245
 QY 160 TG-----GCTGTGCT 183
 DB 246 TGACAGTCT 305
 QY 184 CTGTGCGCTTGTCT 243
 DB 306 CTGTGCGCTTGTCT 365
 QY 244 CCGGCGCAG 303

DB 366 CCGGCGC-----GACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422
 QY 304 CTGTGCGCAG 336
 DB 423 CTGTGCGCAG 455

RESULT 13

AAC85555
 ID AAC85555 standard; cDNA; 526 BP.

XX AAC85555;
 AC

DT 04-JUN-2001 (first entry)
 XX

DE cDNA encoding CDIFP-14, Incyte ID No. 3432292GB1.

XX Human; cell differentiation; CDIFP; agonist; antagonist; epilepsy;
 KW cell proliferation; Alzheimer's disease; schizophrenia disorder;
 KW arteriosclerosis; cancer; atherosclerosis; diabetes mellitus; ss.
 XX Homo sapiens.

XX Key

XX Location/Qualifiers
 FT CDS 54..446
 FT /tag= a
 FT /product= "CDIFP-14"

XX W0200119860-A2.
 XX 22-MAR-2001.
 XX 14-SEP-2000; 2000WO-US25435.

XX 15-SEP-1999; 99US-0154140.
 PR 06-DEC-1999; 99US-0169155.
 XX
 PA (INCY-) INCYTE GENOMICS INC.

PI Tang YT, Hillman JL, Yue H, Reddy R, Lal P, Shah P, Azimzai Y;
 PI Baughn WR, Lu DM, Bardman O, Shin JL, Patterson C;
 XX
 DR WPI; 2001-211447/21.
 DR P-PSDB; AAB47133.

XX Isolated polypeptides and polynucleotides involved in cell
 PT differentiation are used for treatment, prevention and diagnosis of
 PT cell proliferative, developmental and neurological disorders e.g.
 PT cancer and Alzheimer's disease -
 XX

PS Claim 5; Page 127-128; 137pp; English.

XX The sequences given in AAC85545-72 encode human polypeptides involved
 CC in cell differentiation (CDIFP). CDIFP polypeptides and agonists of
 CC these are used to treat a disease or condition associated with
 CC decreased expression of functional CDIFP. An antagonist of CDIFP is
 CC used to treat a disease or condition associated with over expression
 CC of functional CDIFP. CDIFP polypeptides may be used for the treatment,
 CC prevention and diagnosis of cell proliferative, developmental and
 CC neurological disorders, such as Alzheimer's disease, schizophrenia
 CC disorders, arteriosclerosis, cancer, atherosclerosis, diabetes mellitus
 CC and epilepsy. The CDIFP-14 sequence is homologous to human skin
 CC specific protein.
 XX

SQ Sequence 526 BP; 96 A; 176 C; 139 G; 115 T; 0 other;

Query Match 12.0%; Score 72.4; DB 22; Length 526;
 Best Local Similarity 59.0%; Pred. No. 1.2e-09;
 Matches 209; Conservative 0; Mismatches 96; Indels 49; Gaps 3;

QY 30 ACCTGACAACTCAACTCTCTGCGCAGAGTCTCTGCGCAGAGAGAGAGAGAGAGAGAG 89

CC	cancaissimale during a total intra-abdominal colectomy, appendectomy,
CC	and permanent ileostomy. Recombinant vectors comprising Hsp1 cDNA are
CC	introduced into host cells for protein expression. The Hsp1 proteins are
CC	useful for the treatment of epithelial disorders, including dysidirotic
CC	eczema and scabies, cell proliferative disorders including actinic
CC	keratosis and arteriosclerosis, and autoimmune/inflammatory disorders
CC	like acquired immune deficiency syndrome (AIDS) and Addison's disease.
CC	Pharmaceutical compositions comprising Hsp1 proteins are useful for
CC	treating disorders associated with altered Hsp1 expression.
XX	
5Q	Sequence 528 BP; 96 A; 175 C; 131 G; 126 T; 0 other;
	Query Match 11.8%; Score 71; DB 21; Length 528;
	Best Local Similarity 66.9%; Pred. NO. 2.9e-09;
	Matches 117; Conservative 0; Mismatches 55; Indels 3; Gaps 1
QY	162 GCGTGGCCCAAGCTCTTGAGGCTGTAGCCCTAGTCCGAGGGCGGCTCTTCTGAAC 221
	164 GCGTGTGCTCCAGCTCTTGAGGCTGTAGCCCTAGTCCGAGGGCTGTGCTCTCTTGAGCC 222
DB	222 ACCACAGGCGCCACCAACGATGCGGCGCCAGAGGCCCACTCTTGAGCAGGGGACAGTG 281

[illegible]

RESULT 15
ABQ61174
ID ABQ61174 standard; cDNA: 678 BP.

AC ABQ61174;
XX
XX 26-FEB-2003 (first entry)
XX
DE Human epidermal protein-2 encoding sequence.
XX

cytostatic; anti-inflammatory; gene therapy; nutritional supplement

XX Homo sapiens.
KM amyotrophic lateral sclerosis; autoimmune disorder; inflammation; vulnerability; gene; ss.
KM Alzheimer's disease; Huntington's disease;
KM Down, Paul; liver; autoimmune disorder; inflammation;
KM Wound, Paul; liver; autoimmune disorder; inflammation;
SS

AA WO200231111-A2
PN

18-APR-2002

11-OCT-2001: 2001WO-TTS27760

12-OCT-2000: 2000TTS-0687527

XX
PA (HYSE-) HYSEN INC

XX	Tana	VT	Tina
PI			C

Xue AJ, Yang Y, Wehrman T, Drmanac RT; PI XY

DR WPI; 2002-426278/45.
DR N-PCDB; APB43930

DT XX New not, most of the

sources or supplements, or in gene therapy, particularly for treating wounds. Although the use of stem cells in regenerative medicine is still in the early stages, it is an area of active research and development.

PT inflammation -

PS- Claim 1; SEQ ID # 387; 357pp + sequence listing; English

CC The invention relates to 446 newly isolated polynucleotide sequences

CC vulnerable, neuroprotective, immunomodulator, cytostatic and

CC anti-inflammatory. Compositions comprising nucleic acids of the invention
CC are useful for treating a mammalian subject, or as nutritional sources or
CC supplements. These are useful in gene therapy, particularly for treating
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC inflammation. The nucleic acids and polypeptides are also useful in
CC diagnostic and research methods. The sequences given in records
CC AB060788-AB061233 represent polynucleotides of the invention.
CC NOTE: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIFO
CC at http://wipo.int/pub/published_pcf_sequences.

XX
SQ Sequence 678 BP; 129 A; 238 C; 183 G; 128 T; 0 other;

Query Match 11.4%; Score 68.8; DB 24; Length 678;
Best Local Similarity 53.0%; Pred No. 1.2e-08;

Matches 176; Conservative 0; Mismatches 147; Indels 9; Gaps 1;

QY 42 CAATCTCTGCAAGATGTCTGCGACAGAAACAGCAGAGTGCACACCCCAACCAAGT 101
DB 47 CACCCACCCGCGAGATGTCTCTGCGACAGCAGCAGCAGCAGTGCACGCCCCCTCCCAAGT 106
QY 102 GTCCCTCAACCAAGTGTCTGCGACAGAAAGCCAGTACAGTGTCTGCTCCAGCTTCCCTG 161
DB 107 GCAACCCCAAGTGTCTGCGACAGTGTCTGCGACAGTGTCTGCGACAGTGTCTGCGACAGT 166
QY 162 GCTGTGCCCCCAAGCTGTGAGGTCTGTGAGCCCTAGCTCCAGAGGCGGCT-----GCT 212
DB 167 AGTGCCCTCAAGTGTCTGCTGCTGCGACAGTGTCTGCGAGGCTGCTGTGAGCTCCAGCT 226
QY 213 TCTTGAAACCAACAGAGGCGCCACACCAAGTGTCTGCGACAGTGTCTGCGACAGTGTCTGCGAC 272
DB 227 CTGGGGGCGGCTGCTGCTGCGAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 286
QY 273 GAGGCGAGTGTGAGCAGAGGCGGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 332
DB 287 GGTGCCACCGTCAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 346
QY 333 GATCCGATCTGATGCTGAGACAGCGATCT 364
DB 347 GCTGCTGTGAGGAGGAGGAGGCGGCGACACTCT 378

Search completed: February 16, 2004, 06:22:16
Job time : 330 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 05:39:35 ; Search time 2280 Seconds

(without alignments)
6427.894 Million cell updates/sec

Title: US-09-744-315-7

Sequence: 603
1 ccgagacagccttcctcctgc.....aaagataagcttcctgc 603

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_estbm:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estp1:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_iny:*
19: em_gss_pin:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pbg:*
27: em_gss_vit1:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	468	77.6	477	13	BQ300994 PMO-KT004
C 2	409.8	68.0	528	9	AM136771 UI-R-B11-
C 3	398.8	65.1	514	9	AM205844 UI-R-B11-
C 4	321.4	53.3	470	12	B1670520 SPRL 6b H

5	317.4	52.6	470	12	B1670519	B1670519 SPRL 6a H
C 6	286	47.4	407	9	AM137411	AM137411 UI-R-B11-
C 7	189.6	31.4	567	12	BM389883	BM389883 UI-R-CN1-
C 8	186	30.8	485	14	BY708906	BY708906 BY708906
C 9	185	30.8	570	11	AK009036	AK009036 Mus muscu
C 10	184.4	30.7	231	10	BP354854	BP354854 MR2-HT078
C 11	185	30.6	757	14	CB575489	CB575489 AGENCOURT
C 12	182.4	30.2	563	11	AK008201	AK008201 Mus muscu
C 13	182.4	30.2	564	11	AK008001	AK008001 Mus muscu
C 14	182.4	30.2	944	14	CB590520	CB590520 AGENCOURT
C 15	182.2	30.2	543	12	BI277681	BI277681 UI-R-C20-
C 16	175.6	29.1	429	14	W34356	W34356 ma9a09.x1
C 17	174.6	29.0	535	14	CB575977	CB575977 AGENCOURT
C 18	172	28.5	627	14	CB589787	CB589787 AGENCOURT
C 19	171.2	28.4	401	9	AA059561	AA059561 m65f03.r
C 20	171.2	28.4	537	4	BM523362	BM523362 RZPD Mus
C 21	171.2	28.4	638	14	CB588625	CB588625 AGENCOURT
C 22	169.2	28.1	624	14	CB321483	CB321483 AGENCOURT
C 23	166.4	27.6	606	14	CB575565	CB575565 AGENCOURT
C 24	165.4	27.4	552	12	BI285481	BI285481 UI-R-DD0-
C 25	161.2	26.7	601	9	AV071932	AV071932 AV071932
C 26	153.4	25.4	347	14	W18301	W18301 mb87f09.x1
C 27	148.6	24.6	203	12	BI670516	BI670516 SPRL 3a H
C 28	142.6	23.6	307	10	BF545833	BF545833 UI-R-C2-m
C 29	142	23.5	227	10	BE717933	BE717933 MR2-HT078
C 30	138.6	23.0	487	12	BI286064	BI286064 UI-R-DD0-
C 31	133.8	22.2	477	10	BF417300	BF417300 UI-R-CNO-
C 32	119.4	19.8	401	14	W37034	W37034 mb72b10.x1
C 33	111.8	18.5	432	12	BM384766	BM384766 UI-R-CN1-
C 34	111.4	18.5	394	12	BI670517	BI670517 SPRL 4a H
C 35	109.8	17.2	252	14	W11697	W11697 mb21a07.x1
C 36	103.6	16.6	571	12	BI670518	BI670518 SPRL 5a H
C 37	98.6	16.4	282	10	BB565509	BB565509 BB565509
C 38	96.8	16.1	460	11	AK009081	AK009081 Mus muscu
C 39	96.8	16.1	288	13	BY355882	BY355882 BY355882
C 40	96.6	16.0	220	28	BH041328	BH041328 RPCI-24-3
C 41	95.6	15.9	540	9	AW291010	AW291010 UI-R-B12-
C 42	95.6	15.9	832	10	BG198345	BG198345 RGT17603
C 43	92	15.3	271	10	BB565495	BB565495 BB565495
C 44	90.4	15.0	258	10	BB565531	BB565531 BB565531
C 45	89	14.8	248	10	BB565379	BB565379 BB565379

ALIGNMENTS

RESULT 1
BQ300994/c

LOCUS PMO-KT0041-030201-003-b12 KT0041 Homo sapiens cDNA, mRNA sequence. 477 bp mRNA linear EST 16-MAY-2002

DEFINITION BQ300994
VERSION BQ300994.1 GI:20816516

KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 477)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsumura, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jorgensen, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Research
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Qy	160	TGCTGTGCCCCAAGACTCTGTGGGATCTGTGGCCCTTAGCTCCAGAGGCGGCTCTCTCGAA	219
Db	468	TGGCTGTGCCCAAGCTCTGTGGGGCTGTGGCCCTTAGCTCCAGAGGCGGCTCTCTCGAA	409
Qy	220	CCACCAAGGCGCCACACCCGATGCCGCGCGCCAGAGGCCAACCTCTGTGACAGGGGAC	279
Ds	408	CCACCCAGGCGCCACACCCGATGCCGCGCGCCAGAGGCCAACCTCTGTGACAGGGGAC	349
Qy	280	TGGTCAGCAAGGCGGGGGGCTCTGGGCTGTGCGACAGGTTCTGAGGGGCTGTGTCGATCCAG	339
Ds	348	TGGTCAGCAAGGCGGGGGGCTCTGGGCTGTGCGACATGGCTCTGTGAGGCTGTGCTGACTGG	289
Qy	340	ATCCTGATGCTGAGACAGAGCATCTTTGAGGAAAACAAGATCCCAAGGCCCAAGACA	399
Ds	288	ATCCTGATGTTTGAGACAGAGCAT-TTGGAGAAAACAAGATCCCAAGGCCCAAGAAA	230
Qy	400	GCCCCATGTGAGCGCATGGCTTCCCATATACCTCTTTGACTTTTCAAGGCTGAGCTGGA	459
Ds	229	CCCCATTTGATGTCAAGTAGTGTCCAGATACCTCTCTGTGGCTTTTCAAGGCTGAGCTGAG	170
Qy	460	GATTTCCTGTGGGGGATCTGAGCTCTCCCGAAGAGCACTTCTGTGTTTATGACAGA	519
Ds	169	GATTTCCTGTGGAAGGTCTGAGCTCTCCCGAAGAGCACTTCTGTGTTTATGACAGA	110
Qy	520	TGTCAATATGTCCCCCTAAGCCCTGTACCTGACCAAGGATTTGGCAAGTCTTGTGCCCCAATTC	579
Ds	109	TGTCACATGTTCCTTACCCCTGTACCTGTGAAAGATTTGGCAAGTCTTGTGCCCCAATTC	50
Qy	580	GTAAGAAAGATAAGTTC	597
Ds	49	ATCAAAATAATAAGTTC	32

```

RESULT 3
AM205844/c
LOCUS
DEFINITION
AM205844
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
NIH-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/ILNLT at:
www-bio.lnlnl.gov/bbnp/image/image.html
Seq primers: M13 Forward
POLY(A)yes.
Location/Qualifiers
1. 514

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2722901"
/lab_host="RH108 (Life Technologies)"
/clone_idb="NCI CGAP Sub3"
/name="vector: p773D-Pac (Pharmacia) with a modified
polymerase: Site_1; Not I; Site_2; Eco RI; The
NCI CGAP Sub3 library is a subcloned library derived from

```

Query Match	Best Local Similarity	Matches	438; Conservative	66.1%; Score 398.8; DB 9; Length 514; Freq. No. 1.1e-94; Indels 1; Gaps 1.
111	CCAAAGTGTCCCCCAAGAAGCCCAAGTAAAGTGTGCTCCAGCTTCTTGCTGCTGACC	170		
514	CCAAAGTGTCCCCCAAGAAGCCCAAGTAAAGTGTGCTCCAGCTTCTTGCTGCTGACC	455		
171	CAAGCTCTGGGGTCTGTGGCCCTTAGCTCCGAGGGCGGCTGCTTCTTGAAACCAACAGC	230		
454	GACAGCTTGGGGGCTGTGGCCCTTAGCTCCGAGGGCGGCTGCTTCTTGAAACCAACAGC	395		
231	GCCACCAACCGATGCGGGCGCCAGAGGGCCCAACTCTGTGACAGAGGGCACTGGTCAAG	290		
394	GCCACCAACCGATGCGGGCGCCAGAGGGCCCAACTCTGTGACAGAGGGCACTGGTCAAG	335		
291	GCGGGGGGCTTGTGGCTGTGCTCCACGCTTCTTGGGGGCTGTGCTGTGATCAATCTGATGCT	350		
334	GCGGGGGGCTTGTGGCTGTGCTCCACGCTTCTTGGGGGCTGTGCTGTGATCAATCTGATGCT	275		
351	GAGACAGCGATCTTTGAGAGGAAACAAGATCCCAAGAGGGCCCAAGAACAGCCCATCTGA	410		
274	GAGACAGCGAT-TTTGAAGAGAAACAAGATCCCAAGAGGGCCCAAGAACAGCCCATCTGA	216		
411	CGAGTCCCTTCCCAATACCTCTTCTGACTTTCAACAGGTGAGCTGAGAGTTTCTGCT	470		
215	TGATATAGTTCCCAATACCTCTTCTGACTTTCAACAGGTGAGCTGAGAGTTTCTGCT	156		
471	GAGGAGTCTGAGCTCTCCCAAGAGGACCTTTGTTTATATGTAAGATGTCAATATGTC	530		
155	GGAAGGTCTGAAGCTCTCCCAAGAGGACCTTCTGCTGTGATGTCAGAGATGTCATGTT	96		
531	CCCTTACCCCTTACCTGTGAAGAGTTGGCAGTGTGCTTGGCCCAACCTGTGTAATAAGT	590		
95	CCCTTACCCCTTACCTGTGAAGAGTTGGCAGTGTGCTTGGCCCAACCTGTGTAATAATA	36		
591	AAAGTTC	597		

I.M.A.G.E. Consortium/ILNU at:
www.bio.lnln.gov/bbrp/image/image.html
Seq primer: M13 forward
POLYA=yes

FEATURES

source

Location/Qualifiers

1..407

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2715360"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pRT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The NCI CGAP Subj library is a subtracted library derived from the NCI CGAP Subj library, which is a subtracted library derived from Bi. Bi constitutes a mixture of 21 normalized or subtracted NCI CGAP libraries: NCI CGAP_C04, NCI CGAP_P122, NCI CGAP_P128, NCI CGAP_C010, NCI CGAP_C016, NCI CGAP_K1d5, NCI CGAP_K1d12, NCI CGAP_K1d3, NCI CGAP_K1d11, NCI CGAP_Lym2, NCI CGAP_Br2, NCI CGAP_C08, NCI CGAP_C111, NCI CGAP_Lc12, NCI CGAP_Bm23, NCI CGAP_Lu5, NCI CGAP_Lu24, NCI CGAP_Lu19, NCI CGAP_Gc4, NCI CGAP_Gc6, NCI CGAP_Bm25. These 21 libraries were pooled and a single-stranded DNA preparation of the resulting mixture was used as a tracer in a subtractive hybridization with a driver whose composition is detailed below:

NCI CGAP_K1d3 pool 1 LLM 3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids 132376-132391, 145608-145675, 150052-150285); NCI CGAP_K1d5 pool 1, LLM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids 132391-132581, 1471368-1472903, 1492104-1493255); NCI CGAP_Lu5 pool 1, LLM 3575-3582, 3651-3854 (IMAGE Clonoids 1414920-1417991, 1520904-1522439); NCI CGAP_Gc4 pool 1 LLM 3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631, 1469064-1470983, 1475592-1476743); NCI CGAP_P122 pool 1 LLM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759, 1101193-1101959, 1217928-1220615); NCI CGAP_C010 pool 1 LLM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351). Subtraction was performed as previously described [Bonaldi, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6: 791-806.

6. 791-806.
TAG_LIB=NCI CGAP_Lu19
TAG_TISSUE=lung
TAG_SEQ=GACAGC

BASE COUNT

87 a 107 c 113 g 100 t

ORIGIN

Query Match 47.4%; Score 286; DB 9; Length 407;
Best Local Similarity 87.6%; Pred. No. 7e-65;
Matches 324; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

228 GGGGCCACCAAGGATCCGGGCGCAGAGGCCCAACTCTGTGACAGGGGCGATGCTTACG 287
Db 397 GGGGCCACCAAGGATCCGGGCGCAGAGGCCCAACTCTGTGACAGGGGCGATGCTTACG 338
288 AAGCGGGGCTCTGCTCTGCTGCAAGGTTTGGGGGCTGCTGCTGATCAGATCTGAT 347
Db 337 AAGCGGGGCTCTGCTCTGCTGCAAGGTTTGGGGGCTGCTGCTGATCAGATCTGAT 278
348 GCTGACACAGGATCTTTGAGAGAAACAAGATCCCAAGAGGCCCAAGACAGCCCATC 407
Db 277 GCTGACACAGGATCTTTGAGAGAAACAAGATCCCAAGAGGCCCAAGACAGCCCATC 219
408 TGAAGATGCTTCCATATACCTCTTGTGATTTCAAGGCTAGCTGAGGTTTTC 467
Db 218 TGAAGATGCTTCCATATACCTCTTGTGATTTCAAGGCTAGCTGAGGTTTTC 159
468 TGTGGGAGATCTAGCTTCTCCCAAGAGGACCTTCTGTTTATGTACGATGTATAT 527

Db 158 TGTGAGAGTCTGAGCTTCTCCCAAGAGGACCTTCTGCTATGTTCAGATGTACAT 99
QY 528 GTCCCTTACCCCTGTGATCTGACAGGATTTGAGGATGCTTGTGCCCACTCTTAAAAA 587
Db 98 GTCCCTTACCCCTGTGATCTGACAGGATTTGAGGATGCTTGTGCCCACTCTTAAAAA 39
QY 588 GATTAAGTTC 597
Db 38 ATTAAGTTC 29

RESULT 7
LOCUS BM389983/c
DEFINITION UT-R-CNI-cjo-g-12-0-UI s1 UT-R-CNI Rattus norvegicus cDNA clone
ACCESSION UT-R-CNI-cjo-g-12-0-UI 3', mRNA sequence.
VERSION BM389983
KEYWORDS BM389983.1 GI:18190036
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
EST. BM389983
Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS 1 (bases 1 to 567)
TITLE Bonaldi,M.F., Lennon,G. and Soares,M.B.
JOURNAL Normalization and subtraction: two approaches to facilitate gene
MEDLINE discovery
PUBMED Genome Res. 6 (9), 791-806 (1996)
97044477
8889548

COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBR, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bent-soares@uiowa.edu
The sequence contained an oligo-dt track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonaldi poly A
tail. cDNA library preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 forward
POLYA=yes

FEATURES

source

Location/Qualifiers

1..567

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UT-R-CNI-cjo-g-12-0-UI"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pRT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The UT-R-CNI library is a subtracted library derived from the following pool of seven normalized rat libraries: normalized rat seminal vesicles, normalized rat penis, normalized rat bladder, normalized rat cervix, and normalized rat brown adipose, normalized rat fundus, and normalized rat salivary gland. It was constructed according to the procedure described by Bonaldi, Lennon & Soares (Genome Research 6: 791-806, 1996). For construction of the CNI library, plasmid DNA from the pool of normalized libraries was electroporated into competent bacteria for the production of single-stranded circular DNA. This was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a pool of about 34,000 clones from the Rat Unigene Set corresponding to plates R-5-AA-NN excluding plates R-5-MM and MN. This pool represented 40%


```

Query Match 30.2%; Score 182.4; DB 11; Length 564;
Best Local Similarity 76.0%; Pred. No. 2e-37;
Matches 310; Conservative 0; Mismatches 76; Indels 22; Gaps 6

QY 14 CTCTGCCTCTCTGTGCACCTGGACAACCTAACCTCTCCGCAAGATGTCTGCAGCAGAAC 73
Db 25 CTTCTGTCTCTTTGAATCTAGTCAACT-GACTCTGCGCAAGATGTCTGTCCAGCAGAGC 83
QY 74 CAGCAGCAGTGGCAACCCCAACCCAAAGTCTCTTCAACCAAGTGTCCCCA----- 124
Db 84 CAGAAAGCATGTCCAACTTCTCTCCCAAGTCTCTTCCCAAAGTCTCTCCCAAAGTCCCC 143
QY 125 --AAAGGCCAGTACAGTGTCTGCTCAGCTTCCGTGTGGCTGTGCCCCAAGCTCTGGG 181
Db 144 AGGAAGAGCAGACACAGTGTCTGGCTGAGGCTCTTCTGCTGTGTCTACAACTCTGGG 203
QY 182 GTCT---GTGGGCTTAAGCTTCCAGAGGCGGGCTGCTTCTTGAACCAACCAAGGCG-- 235
Db 204 GGCTCAAGTCTTCCAGATCTTAGGAGAGGCTCTGCTGTGAGCACCAACAGGCGCAGATCC 263
QY 236 CACCATGCGCGCGCCAGAGGCGCCAACTCTGTGA CAGGGGAGAGTGTACGCAAGCGGG 295
Db 264 CACAAATGAGGCGCAGAGAGCTCCAGTTCCTGTGACGTGGCAGTGTGAGCAAGTCTGGG 323
QY 236 GGCTTGGCTGCTGCCACGATTCTTGGGGGCTGCTGTGATCCAGATCTGTATGCTGAAG 355
Db 324 GGCTCAGGCTGTGGCCAGAGCTCTGGGGACCTGCTGTGACCTGGAGCATGATGTGAAGAC 383
QY 356 AAGCATCTTTGGAGAAACAAGATTCACAGAGGCGCAAGAACACACC 403
Db 384 AAGACA-GTCTGGAGAAATCAACA--TCCAGAGGATCAAGAAAACCCC 428

```

FEATURES
source

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30290065"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIM MGC 136"
/insect_vector="pCMV-Sport6.1.cdbb; Site 1: EcoRV; Site 2
NotI; Normalized, full-length enriched library from pool
of mouse embryonic limb, maxilla and mandible, embryonic

```

day 17.5, 18.5 and newborn (mable (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACCTGTTGAGATCGGAGCGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Intitrogen Corp. Note: this is a NIH_MGC Library."

Query Match	30.2%	Score 182.4	DB 14	Length 944
Best Local Similarity	76.0%	Pred. No. 2.3e-37		
Matches 310	Conservative 0	Mismatches 76	Indels 22	Gaps 6
QY	14	CTCTGCTCTCTTGGACCTTGACAACTCAACTCTTGCCAAAGATGTGCTGCGACAGAAC	73	
Db	6	CTCTGCTCTCTTTGAATCTAGTCAACT-GACTCTTGCCAAAGATGTCTGCGACAGAAC	64	
QY	74	CAGCAGCAGTGGCAACCCCAACCAATGTCTCTCAACCCAAAGTGTCCCCA-----	124	
Db	65	CAGAGCAGTGGCAACCTCTCCCAATGTCTCTCTCCAAAGTGTCTCCCAAGTCCCC	124	
QY	125	---AAGAGCCAGTACAGATGTCTGCTCCAGTCTCTGCTGCTGCTGCCCAAGCTCTGGG	181	
Db	125	AAGAAAGAGACAGCAGATGTCTGCTGCGACCTCTTCTGCTGTGCTCAAACTCTGGG	184	
QY	182	GTCT--GTGACCTTGACTCCAGAGGCGTGTCTTCTCTGAACCAACACAGGCGC---CAC	235	
Db	185	GACTCAATCTTCCCAATCTGAAGAGGCTGTGCTCTAGGCACACAGGCGCAGTCTC	244	
QY	236	CACCGATCCCGCGCCAGAGGCCCAACTCTGTGAACAGGGGCGAGTGTAGCAAGCGGG	295	
Db	245	CACCAATCCAGGCCAGAGTCTCAATCTCTGTGAACCGTGGCAGTGTGTAGCAATCTTGGG	304	
QY	296	GACTTGTGCTGTGCCACAGCTTCTGGGGGCTGTGCTGAATCCAGATCTTGATCTGAAC	355	
Db	305	GGGTGAGGCTGTGGCCACAGCTCTGGGGAACTGTGCTGAATCTGGAGATGATGTGGAAC	364	
QY	356	AAGCATCTTTGAGAGAAACAAAGATCCCAAGAGGCCAAGAACAGCCC	403	
Db	365	AAGACA-GTCTGAGAAATCAACA-TCCCAAGGGGTCAAGAAAGGCC	409	

RESULT 15	BI277681/c	543 bp	mRNA	linear	EST 19-JUL-2001
LOCUS	BI277681				
DEFINITION	UI-R-CCO-byf-b-12-0-UI.s1 UI-R-CCO Rattus norvegicus cDNA clone				
ACCESSION	BI277681				
VERSION	BI277681.1	GI:14923791			
KEYWORDS	EST.				
SOURCE	Rattus norvegicus (Norway rat)				
ORGANISM	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE	1 (bases 1 to 543)				
AUTHORS	Bonaldi,M.F., Lennon,G. and Soares,M.B.				
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery				
JOURNAL	Genome Res. 6 (9), 791-806 (1996)				
MEDLINE	97044477				
PubMed	8889548				
COMMENT	Contact: Soares, MB				

Coordinate Laboratory for Computational Genomics
University of Iowa
375 Newton Road / 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-scotars@iowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the non-normalized penis library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLY(A)yes.

FEATURES
Source

Location/Qualifiers
1..543
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C20-b-12-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-C20"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C20 library is a non-normalized library constructed from rat penis tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bomalio, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG LIB=UI-R-C20
TAG TISSUE=penis
TAG_SEQ=TTGCGGACACA"
BASE COUNT 118 a 144 c 162 g 119 t
ORIGIN

Query Match 30.2%; Score 182.2; DB 12; Length 543;

Best Local Similarity 76.4%; Pred. No. 2.2e-37;

Matches 282; Conservative 0; Mismatches 68; Indels 19; Gaps 4;

QY	46	TTCTGCGCAAGATGCTGCGCAGACGAGCAACGAGCCCAACCCCAAGATGTC	105
DB	527	TCATGCCAAGATGCTGCGCAGACGAGCAACGAGCAAGCTCTCCCAAGTCCC	468
QY	106	CTCAACC-----AAGTGTCCCAAGAGCCCAAGTGTCTGCTCCAGC	153
DB	467	CTCCCCAAGTGTCTCCCAAGTGTCCCAAGAGCAAGTGTCTGCTCCAGC	408
QY	154	TTCTGCGCTGAGCCCAAGCTCTGAGGCTCT-----GTGGCCCTAGCTCCAGAGGCGGCTG	210
DB	407	TTCTTCTGCTGCTGCTCAAGCTCTGAGGCTCTCAAGCTCTGAGGAGGCTG	348
QY	211	CTTCTGAGCAACCAAGCGC--CAACCAAGATGCGCGCCAGAGGCCCACTCTG	267
DB	347	CTGCTGAGCCACCAAGCGCGAGTCCCAAGATGAGCGCGAGAGCTCCAGTTCCTG	288
QY	268	TGACAGGAGGAGTGTGAGCAAGCGGAGGCTCTGCTGCTGCAAGTTCTGAGGAGCTG	327
DB	287	TGACCTGAGAGTGTGAGCAAGTGTGAGGAGCTCTGCTGCTGCAAGTTCTGAGGAGCTG	228
QY	328	CTGCTGATCCAGATCTGATGTAAGCAAGCATCTTTGAGAGAAACAAATCCCAAG	387
DB	227	CTGCTAAGCTGAGAGATGATGTAAGCAATGAGA-CTCTGAGAGAAAGCATATCCCAAG	169
QY	388	AGGCCAAGA 396	
DB	168	GCTCAAGGA 160	

Search completed: February 16, 2004, 07:54:52
Job time : 2284 secs

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 02:57:33 ; Search time 3257 Seconds

7573.996 Million cell updates/sec

Title: US-09-744-315-7

Sequence: 1 caggaacagcctctcctgc.....aaaagatcaagttccgctgc 603

Scoring table: IDENTITY_NUC

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

```
Minimum DB seq length: 0
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

1:  gb_ba:*
2:  gb_hrg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hrg_hum:*
31: em_hrg_in:*
32: em_hrg_other:*
33: em_hrg_mus:*
34: em_hrg_pln:*
35: em_hrg_rod:*
36: em_hrg_mam:*
37: em_hrg_vir:*
38: em_sy:*
39: em_hrgo_hum:*
40: em_hrgo_mus:*
41: em_hrgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	100	100	1	...
2	95	95	95	2	...
3	90	90	90	3	...
4	85	85	85	4	...
5	80	80	80	5	...
6	75	75	75	6	...
7	70	70	70	7	...
8	65	65	65	8	...
9	60	60	60	9	...
10	55	55	55	10	...
11	50	50	50	11	...
12	45	45	45	12	...
13	40	40	40	13	...
14	35	35	35	14	...
15	30	30	30	15	...
16	25	25	25	16	...
17	20	20	20	17	...
18	15	15	15	18	...
19	10	10	10	19	...
20	5	5	5	20	...

C	1	561.2	93.1	52355	9	ALJ56426	ALJ56426 Human DNA
C	2	511.4	84.8	617	6	BDI81469	BDI81469 A gene of
C	3	511.4	84.8	617	9	ABO48288	ABO48288 Homo sapi
C	4	480.4	79.7	76369	2	ACO23208	ACO23208 Homo sapi
C	5	465.4	77.2	76369	2	ACO23208	ACO23208 Homo sapi
C	6	360.4	59.8	52355	9	ALJ56426	ALJ56426 Human DNA
C	7	360.4	59.8	117084	9	ALJ39247	ALJ39247 Human DNA
C	8	198	32.8	228396	2	ACI40541	ACI40541 Macaca mu
C	9	187.6	31.1	199220	2	ACI32274	ACI32274 Mus muscu
C	10	186.2	30.9	2275	6	BDI81458	BDI81458 A gene of
C	11	185.4	30.7	195503	2	ACI37851	ACI37851 Rattus no
C	12	185.4	30.7	261316	2	ACO95371	ACO95371 Rattus no
C	13	185.4	30.7	347801	2	ACI18786	ACI18786 Rattus no
C	14	184.4	30.6	566	10	AFI76515	AFI76515 Mus muscu
C	15	184	30.5	117084	9	ALJ39247	ALJ39247 Human DNA
C	16	183.8	30.5	195503	2	ACI37251	ACI37251 Rattus no
C	17	183.8	30.5	311774	2	ACI18873	ACI18873 Rattus no
C	18	183.8	30.5	347801	2	ACI18786	ACI18786 Rattus no
C	19	181	30.0	33339	10	AFB03430	AFB03430 Mos muscu
C	20	177.2	29.4	199220	2	ACI32274	ACI32274 Mus muscu
C	21	169.4	28.1	213540	2	ACI35289	ACI35289 Mus muscu
C	22	167.2	27.7	311774	2	ACI18873	ACI18873 Rattus no
C	23	142.6	23.6	689	6	AXS25830	AXS25830 Sequence
C	24	142.6	23.6	261316	2	ACO95371	ACO95371 Rattus no
C	25	134.4	22.3	148	6	AXK93799	AXK93799 Sequence
C	26	103.6	17.2	188285	9	ALJ35842	ALJ35842 Human DNA
C	27	82.8	13.7	96013	9	ALJ53779	ALJ53779 Human DNA
C	28	72.6	12.0	619	9	HMMZD48C9	AF086289 Homo sapi
C	29	72.4	12.0	526	6	AY100241	Sequence
C	30	72	11.9	625	6	AF050580	AF050580 Homo sapi
C	31	68.8	11.4	116067	9	ALI62596	ALI62596 Human DNA
C	32	68.6	11.4	265338	2	ACI27247	ACI27247 Mus muscu
C	33	67.2	11.1	231037	2	ACI14064	ACI14064 Rattus no
C	34	67.2	11.1	293137	2	ACI14036	ACI14036 Rattus no
C	35	67	11.1	293137	2	ACI10356	ACI10356 Rattus no
C	36	66.8	11.1	126610	2	ACI32604	ACI32604 Mus muscu
C	37	66.8	11.1	126610	2	ACI32604	ACI32604 Mus muscu
C	38	62.6	10.4	1210	9	BC038391	BC038391 Homo sapi
C	39	62.4	10.3	96013	9	ALJ53779	ALJ53779 Human DNA
C	40	62.4	10.3	25985	2	ACI02992	ACI02992 Rattus no
C	41	59.6	9.3	1397	9	BC031811	BC031811 Homo sapi
C	42	59.2	9.8	188285	9	ALJ5842	ALJ5842 Human DNA
C	43	58	9.6	557	11	G55179	G55179 SHC-100235
C	44	52.2	8.7	544	3	AY113557	AY113557 Drosophil
C	45	51.8	8.6	265338	2	ACI27247	ACI27247 Mus muscu

ALIGNMENTS

RESULT 1	AL356426/c	52355 bp	DNA	linear	PRI 15-NOV-2001
LOCUS	AL356426				
DEFINITION	Human DNA sequence from clone RP1-11013 on chromosome 1q21.1-21.3, complete sequence.				
ACCESSION	AL356426				
VERSION	AL356426.11	GI:16972880			
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Placentalia; Catarrhini; Hominoidea; Homo.				
REFERENCE	1 (bases 1 to 52355)				
AUTHORS	Johnson,C.				
TITLE	Direct Submision				

JOURNAL

Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk
hunquerry@sanger.ac.uk Clonerequest@sanger.ac.uk
 On Nov 16, 2001 this sequence version replaced gi:116030120.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone, and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RPI-11013 is from the library RPI-1 constructed by the group of Pierre de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2
 IMPORTANT: This sequence is not the entire insert of clone RPI-11013. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
 The true left end of clone RPI-5210 is at 50356 in this sequence. The true right end of clone RPI-9195 is at 2000 in this sequence.

FEATURES

source

1..52355
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="q21.1-21.3"
 /clone="RPI-11013"
 /clone_id="RPI-1"

BASE COUNT 15094 a 10374 c 10652 g 16235 t
 ORIGIN

Query Match 93.1%; Score 561.2; DB 9; Length 52355;
 Best Local Similarity 97.8%; Pred. No. 2,7e-124;
 Matches 569; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

19 GCTCTCTGCACTGGAACAATCACTCTCTGCAAGATGCTCTGCCAGCAAGAACGCA 78
 16087 GCTTGTCTTCTCAAGCAATCACTCTCTGCAAGATGCTCTGCCAGCAAGAACGCA 16028
 79 GCAAGTGCACACCCCAACCAAGTCTCTCAACCAAGTCTCTGCCAGCAAGAACGCA 138
 16027 GCAAGTGCACACCCCAACCAAGTCTCTCAACCAAGTCTCTGCCAGCAAGAACGCA 15968
 139 GTGTCTGCTCAAGTCTCTCTGCTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 198
 15967 GTGTCTGCTCAAGTCTCTCTGCTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 15908
 199 CGAGGGCGGCTGCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 258
 15907 CGAGGGCGGCTGCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 15848
 259 CAATCTCTGTAAGAGGCAAGTGTCAAGCAAGGCGGGGCTCTGCTCTGCTCTGCTCTGCT 318
 15847 CAATCTCTGTAAGAGGCAAGTGTCAAGCAAGGCGGGGCTCTGCTCTGCTCTGCTCTGCTCT 15788
 319 TGGGGGGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 378

Db 15787 TGGGGGGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 15728

QY 379 AATCCCAAGAGGCAAGAACAGCCCATCTGACGATGCTTCCATATACCTCTTCTG 438

Db 15727 AATCCCAAGAGGCAAGAACAGCCCATCTGACGATGCTTCCATATACCTCTTCTG 15668

QY 439 ACTTTACAGGCTGACCTGAGGTTTCTCTGCGGGGATCTGAGCTTCTCCCAAGAGCA 498

Db 15667 ACTTTACAGGCTGACCTGAGGTTTCTCTGCGGGGATCTGAGCTTCTCCCAAGAGCA 15608

QY 499 CTCTCTGTTTATGACAGATGATATATGCTCCCTACCTGATACCTGCAAGATG 558

Db 15607 CTCTCTGTTTATGACAGATGATATATGCTCCCTACCTGATACCTGCAAGATG 15548

QY 559 GCAAGTCTTGTGCTCCCACTCTGTAAGAAAGATTAAGTTCTG 600

Db 15547 GCAAGTCTTGTGCTCCCACTCTGTAAGAAAGATTAAGTTCTG 15506

RESULT 2

BD181469

LOCUS BD181469 617 bp DNA linear PAT 15-MAY-2003
 DEFINITION A gene of which expression changes in psoriasis and a method for e
 xamination directed to said gene.

ACCESSION BD181469
 VERSION JP 2002330770-A/17.

KEYWORDS Homo sapiens (human)

SOURCE ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 617)

AUTHORS Imai,Y., Wakimoto,K., Yamada,E., Chiba,H. and Okubo,K.

TITLE A gene of which expression changes in psoriasis and a method for e

JOURNAL examination directed to said gene

PATENT: JP 2002330770-A 17 19-NOV-2002;

TANABE SEIYAKU CO LTD

OS Homo sapiens (human)

PN JP 2002330770-A/17

PF 19-NOV-2002

PI 25-MAY-2001 JP 2001156529

PI YUI IMAI,KOTI WAKIMOTO,ERICO YAMADA,HIROAKI CHIBA,KOSAKU PI

OKUBO PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21 PC

,C12N5/10,C12Q1/68, PC G01N33/53,G01N33/56,C12N15/00,C12N5/00 CC A gene

of which expression changes in psoriasis and a method CC

CC Examination directed to said gene

FT Key Location/Qualifiers

FT CDS Location/Qualifiers

1..617
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

BASE COUNT 132 a 200 c 156 g 129 t

ORIGIN

Query Match 84.8%; Score 511.4; DB 6; Length 617;
 Best Local Similarity 92.1%; Pred. No. 3.9e-112;
 Matches 550; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

QY 1 CAGGAACAGCTTCTCTGCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 60

Db 3 CAGGAACAGCTTCTCTGCTCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 62

QY 61 CTGCCAGAGAACAGAGAGAGTGCACACCCCAAGTGTCCCTCAACCAAGTGTCC 120

Db 63 CTGCCAGAGAACAGAGAGAGTGCACACCCCAAGTGTCCCTCAACCAAGTGTCC 122

QY 121 CCAGAAAGCCAGTACAGTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 180

TITLE
JOURNAL
COMMENT

Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Strange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J.,
Tirelli, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A. and Zody, M.

Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6957762.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L6490

Center clone name: 542_E_22

* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
784 883: contig of 783 bp in length
884 883: gap of 100 bp
1704 1703: contig of 820 bp in length
1804 2625: contig of 822 bp in length
2626 2725: gap of 100 bp
2726 3512: contig of 787 bp in length
3513 4414: contig of 802 bp in length
4415 4514: gap of 100 bp
4515 5306: contig of 792 bp in length
5307 5406: gap of 100 bp
5407 6195: contig of 789 bp in length
6196 6295: gap of 100 bp
6296 7099: contig of 804 bp in length
7100 7199: gap of 100 bp
7200 7998: contig of 799 bp in length
7999 8098: gap of 100 bp
8099 8891: contig of 793 bp in length
8892 8991: gap of 100 bp
8992 9783: contig of 792 bp in length
9784 9883: gap of 100 bp
9884 10701: contig of 818 bp in length
10702 10801: gap of 100 bp
10802 11626: contig of 825 bp in length
11627 11726: gap of 100 bp
11727 12550: contig of 824 bp in length
12551 12650: gap of 100 bp
12651 13429: contig of 779 bp in length
13430 13529: gap of 100 bp
13530 14321: contig of 792 bp in length
14322 14421: gap of 100 bp
14422 15208: contig of 787 bp in length
15209 15308: gap of 100 bp
15309 16072: contig of 764 bp in length
16073 16172: gap of 100 bp
16173 16910: contig of 738 bp in length
16911 17010: gap of 100 bp
17011 17815: contig of 805 bp in length
17816 17915: gap of 100 bp
17916 18725: contig of 810 bp in length

18726 18825: gap of 100 bp
18826 19661: contig of 836 bp in length
19662 19761: gap of 100 bp
19762 20528: contig of 767 bp in length
20529 20628: gap of 100 bp
20629 21438: contig of 810 bp in length
21439 21539: gap of 100 bp
21539 22365: contig of 827 bp in length
22366 22465: gap of 100 bp
22466 23287: contig of 822 bp in length
23288 23387: gap of 100 bp
23388 24152: contig of 765 bp in length
24153 24252: gap of 100 bp
24253 25060: contig of 808 bp in length
25061 25161: gap of 100 bp
25161 25945: contig of 785 bp in length
25946 26045: gap of 100 bp
26046 26827: contig of 782 bp in length
26828 26927: gap of 100 bp
26928 27721: contig of 794 bp in length
27722 27821: gap of 100 bp
27822 28580: contig of 859 bp in length
28581 28780: gap of 100 bp
28781 29619: contig of 839 bp in length
29620 29719: gap of 100 bp
29720 30536: contig of 817 bp in length
30537 30636: gap of 100 bp
30637 31423: contig of 787 bp in length
31424 31523: gap of 100 bp
31524 32324: contig of 801 bp in length
32325 32424: gap of 100 bp
32425 33204: contig of 780 bp in length
33205 33304: gap of 100 bp
33305 34067: contig of 763 bp in length
34068 34167: gap of 100 bp
34168 34884: contig of 817 bp in length
34885 35084: gap of 100 bp
35085 35869: contig of 785 bp in length
35870 36759: contig of 790 bp in length
36760 37659: gap of 100 bp
37660 37759: contig of 800 bp in length
37760 38580: contig of 821 bp in length
38581 38680: gap of 100 bp
38681 39497: contig of 817 bp in length
39498 39597: gap of 100 bp
39598 40384: contig of 787 bp in length
40385 40484: gap of 100 bp
40485 41285: contig of 801 bp in length
41286 41385: gap of 100 bp
41386 42178: contig of 793 bp in length
42179 43062: contig of 784 bp in length
43063 43162: gap of 100 bp
43163 43944: contig of 782 bp in length
43945 44044: gap of 100 bp
44045 44627: contig of 783 bp in length
44628 44927: gap of 100 bp
44928 45728: contig of 801 bp in length
45729 45828: gap of 100 bp
45829 46626: contig of 798 bp in length
46627 46726: gap of 100 bp
46727 47607: contig of 781 bp in length
47608 47607: gap of 100 bp
47608 48432: contig of 825 bp in length
48433 48532: gap of 100 bp
48533 49357: contig of 825 bp in length
49358 49457: gap of 100 bp
49458 50372: contig of 815 bp in length
50373 50372: gap of 100 bp
50373 51159: contig of 787 bp in length
51160 51159: gap of 100 bp

[illegible]

NOTE: This record contains 85 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

1	783: contig of 783 bp in length
784	gap of 100 bp
884	gap of 100 bp
884	1703: contig of 820 bp in length
1704	1803: gap of 100 bp
1804	2625: contig of 822 bp in length
2626	2725: gap of 100 bp
2726	3512: contig of 787 bp in length
3513	3612: gap of 100 bp
3613	4414: contig of 802 bp in length
4415	4514: gap of 100 bp
4515	5306: contig of 792 bp in length
5307	5406: gap of 100 bp


```

* 5407 6195: contig of 789 bp in length
* 6196 6295: gap of 100 bp
* 6296 7098: contig of 804 bp in length
* 7100 7199: gap of 100 bp
* 7200 7998: contig of 799 bp in length
* 8099 8891: contig of 793 bp in length
* 8892 9783: contig of 792 bp in length
* 9784 9883: gap of 100 bp
* 9884 10701: contig of 818 bp in length
* 10702 10801: gap of 100 bp
* 10802 11626: contig of 825 bp in length
* 11627 11725: gap of 100 bp
* 11727 12550: contig of 824 bp in length
* 12551 12650: gap of 100 bp
* 12651 13428: contig of 779 bp in length
* 13430 13529: gap of 100 bp
* 13530 14321: contig of 792 bp in length
* 14322 14421: gap of 100 bp
* 14422 15308: contig of 787 bp in length
* 15309 16072: contig of 764 bp in length
* 16073 16172: gap of 100 bp
* 16173 16910: contig of 738 bp in length
* 16911 17010: gap of 100 bp
* 17011 17815: contig of 805 bp in length
* 17816 17915: gap of 100 bp
* 17916 18725: contig of 810 bp in length
* 18726 18825: gap of 100 bp
* 18826 19661: contig of 836 bp in length
* 19662 19761: gap of 100 bp
* 19762 20528: contig of 767 bp in length
* 20529 21438: gap of 100 bp
* 21439 22365: contig of 810 bp in length
* 22366 22465: gap of 100 bp
* 22466 23287: contig of 822 bp in length
* 23288 24153: contig of 765 bp in length
* 24153 24253: gap of 100 bp
* 24253 25060: contig of 808 bp in length
* 25061 25161: gap of 100 bp
* 25161 25945: contig of 785 bp in length
* 25946 26045: gap of 100 bp
* 26046 26827: contig of 782 bp in length
* 26828 27721: gap of 100 bp
* 27722 27821: contig of 794 bp in length
* 27822 28680: contig of 859 bp in length
* 28681 28780: gap of 100 bp
* 28781 29619: contig of 839 bp in length
* 29620 29719: gap of 100 bp
* 29720 30536: contig of 817 bp in length
* 30537 31423: gap of 100 bp
* 31423 31523: contig of 787 bp in length
* 31524 32324: gap of 100 bp
* 32325 32424: gap of 100 bp
* 32425 33204: contig of 780 bp in length
* 33205 33304: gap of 100 bp
* 33305 34067: contig of 763 bp in length
* 34068 34167: gap of 100 bp
* 34168 34984: contig of 817 bp in length
* 34985 35084: gap of 100 bp
* 35085 35869: contig of 785 bp in length
* 35870 35969: gap of 100 bp
* 35970 36759: contig of 790 bp in length
* 36760 36859: gap of 100 bp
* 36860 37759: contig of 800 bp in length
* 37760 38580: gap of 100 bp
* 38580: contig of 821 bp in length

```

```

* 38581 38680: gap of 100 bp
* 38681 39497: contig of 817 bp in length
* 39498 39597: gap of 100 bp
* 39598 40384: contig of 787 bp in length
* 40385 40484: gap of 100 bp
* 40485 41285: contig of 801 bp in length
* 41286 41385: gap of 100 bp
* 41386 42178: contig of 793 bp in length
* 42179 42278: gap of 100 bp
* 42279 43062: contig of 784 bp in length
* 43063 43162: gap of 100 bp
* 43163 43944: contig of 782 bp in length
* 43945 44044: gap of 100 bp
* 44045 44827: contig of 783 bp in length
* 44828 44927: gap of 100 bp
* 44928 45728: contig of 801 bp in length
* 45729 45828: gap of 100 bp
* 45829 46626: contig of 798 bp in length
* 46627 46726: gap of 100 bp
* 46727 47507: contig of 781 bp in length
* 47508 47607: gap of 100 bp
* 47608 48432: contig of 825 bp in length
* 48433 48532: gap of 100 bp
* 48533 49357: contig of 825 bp in length
* 49358 49457: gap of 100 bp
* 49458 50272: contig of 815 bp in length
* 50273 50372: gap of 100 bp
* 50373 51159: contig of 787 bp in length
* 51160 51259: gap of 100 bp
* 51260 52047: contig of 787 bp in length
* 52047 52146: gap of 100 bp
* 52147 52918: contig of 772 bp in length
* 52919 53018: gap of 100 bp
* 53019 53827: contig of 809 bp in length
* 53828 53927: gap of 100 bp
* 53928 54701: contig of 773 bp in length
* 54701 54800: gap of 100 bp
* 54801 55587: contig of 787 bp in length
* 55588 55687: gap of 100 bp
* 55688 56511: contig of 824 bp in length
* 56512 56612: gap of 100 bp
* 56612 57397: contig of 786 bp in length
* 57398 57497: gap of 100 bp
* 57498 58314: contig of 817 bp in length
* 58315 58414: gap of 100 bp
* 58415 59251: contig of 837 bp in length
* 59252 59351: gap of 100 bp
* 59352 60133: contig of 782 bp in length
* 60134 60233: gap of 100 bp
* 60234 61033: contig of 800 bp in length
* 61034 61133: gap of 100 bp
* 61134 61932: contig of 799 bp in length
* 61933 62032: gap of 100 bp
* 62033 62819: contig of 787 bp in length
* 62820 62919: gap of 100 bp
* 62920 63707: contig of 788 bp in length

Query Match      77.2%; Score 465.4; DB 2; Length 76369;
Best Local Similarity 90.2%; Pred. No. 2.ee-101;
Matches 518; Conservative 0; Mismatches 54; Indels 2; Gaps 2;

```

QY 19 GCGCTCCCTGCACTGAGCAACTCACTCCGCGCAAGTATGCTGCGAGAGAACGCA 78
 Db 28106 GCTTTGTTTTCAGAGCAACTCACTCTGCGCAAGTATGCTGCGAGAGAACGCA 28165

QY 79 GCAAGTGCACCCCAACCAAGTGTCCCTCAACCAAGTGTCCCAAGAGCCCAAGTACA 138
 Db 28166 GCAAGTGCACCCCAACCAAGTGTCCCTCAACCAAGTGTCCCAAGAGCCCAAGTACA 28225

QY 139 GTGCTGCTCAGCTTCCTGCTGCTGCGCCCAAGCTCGGGGTCTGTGGCCCTAGCTC 198
 Db 28226 GTGCTGCTCAGCTTCCTGCTGCTGCGCCCAAGCTCGGGGTCTGTGGCCCTAGCTC 28285

Db	Accession	Version	Source	Organism	Reference Title	Journal	Comment
OY	199	CGAGGGGCGGCTGTCTCCGAACCA	CA	CA	GGGCGCCACCA	CCAGATGCCGGGCCAGAGGGC	258
Db	28286	CGAGGGGCGGCTGTCTCCGAACCA	CA	CA	GGGCGCCACCA	CCAGATGCCGGGCCAGAGGGC	28345
OY	259	CAACTCTGTGACAGGGCAGTGTG	CA	CA	AGGCGGGGCTGTGCTGTGCCACGGTTC	318	
Db	28346	CAACTCTGTGACAGGGCAGTGTG	CA	CA	AGGCGGGGCTGTGCTGTGCCACGGTTC	28405	
OY	319	TGGGGGGCTGCTGTGATCCAGAT	CC	AT	GTGTGAGAACACCATCTTTGAGAGAA	CAG 378	
Db	28406	TGGAGGCTGCTGTGATCCAGAT	CC	AT	GTGTGAGAACACCATCTTTGAGAGAA	CAG 28464	
OY	379	AATCCCAAGAGGCCAAGAACAG	CC	CC	CCATCTGACGATGCTTCCCATATAC	CTTCTG 438	
Db	28465	AATCCCAAGAGGCCAAGAACAG	CC	CC	CCATCTGACGATGCTTCCCATATAC	CTTCTG 28524	
OY	439	ACTTTCACAGGCTGAGCTGAG	GT	TT	CTGTGGGGGATGTAGCTCTCTCCGAAGCA	498	
Db	28525	GCTTTCACAGGCTGAGCTGAG	GT	TT	CTGTGGGGGATGTAGCTCTCTCCGAAGCA	28583	
OY	499	CTTCTGTGTATGTACAGGATG	AT	AT	ATGCCCCACCCCGTATCTGCCAAGATG	558	
Db	28584	CTTCTGTGTATGTACAGGATG	AT	AT	ATGCCCCACCCCGTATCTGCCAAGATG	28643	
OY	559	GCACTGCTGTGTGCCCAACTG	TAT	AAAAAGATTA	592		
Db	28644	GCACTGCTGTGTGTGCCCAACT	TAT	AAAAAGATTA	28677		
RESULT 6							
AL356426							
LOCUS	AL356426	53355 bp	DNA	linear	PR1 15-NOV-2001		
DEFINITION	Human DNA sequence from clone RP1-11013 on chromosome 1q21.1-21.3,						
ACCESSION	AL356426						
VERSION	AL356426.11	GI:16972880					
KEYWORDS	HTG.						
SOURCE	Homo sapiens (human)						
ORGANISM	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.						
REFERENCE	Johnson, C.						
AUTHORS	Direct Submission						
TITLE	Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,						
JOURNAL	CambridgeShare, CB10 1SA, UK. E-mail enquiries:						
	humbug@sanger.ac.uk Clone requests: clonesquest@sanger.ac.uk						
	On Nov 16, 2001 this sequence version replaced gi:16030120.						
	During sequence assembly data is compared from overlapping clones.						
	Where differences are found these are annotated as variations						
	together with a note of the overlapping clone name. Note that the						
	variation annotation may not be found in the sequence submission						
	corresponding to the overlapping clone, as we submit sequences with						
	only a small overlap as described above.						
	This sequence was finished as follows unless otherwise noted: all						
	regions were either double-stranded or sequenced with an alternate						
	chemistry or covered by high quality data (i.e., paired quality >=						
	30); an attempt was made to resolve all sequencing problems, such						
	as compressions and repeats; all regions were covered by at least						
	one plasmid subclone or more than one M13 subclone; and the						
	assembly was confirmed by restriction digest. The following						
	abbreviations are used to associate primary accession numbers given						
	in the feature table with their source databases: Em; EMBL; Sw;						
	SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP						
	database can be found at						
	http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence						
	was generated from part of bacterial clone contigs of human						
	chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping						
	Group. Further information can be found at						
	http://www.sanger.ac.uk/HGP/Chrl						
	RP1-11013 is from the library RPCL-1 constructed by the group of						
	Pleier de Jong. For further details see						
	http://www.chori.org/bacpac/home.htm						

FEATURES	source	1..52355
VECTOR: PCVPC2	/organism="Homo sapiens"	
IMPORTANT: This sequence is not the entire insert of clone RP1-11013. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone RP1-52010 is at 50356 in this sequence. The true right end of clone RP1-9105 is at 2000 in this sequence.	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	/chromosome="1"	
	/map="q21.1-21.3"	
	/clone="RP1-11013"	
	/clone_1kb="PFI-1"	
	/location/Qualifiers	
	1..52355	
BASE COUNT	15094 a 10374 c 10652 g 16235 t	
ORIGIN		
Query Match	59.8%; Score 360.4; DB 9; Length 52355;	
Best Local Similarity	80.8%; Pred. No. 4.4e-76;	
Matches 459; Conservative	0; Mismatches 101; Indels 8; Gaps 3;	
Dy	40	CTCAACCTCCGCAAGATGTCCTGCGACAGAGAAACCGACAGTGCACACCCCAACCCA 99
Db	50558	CTGCATTCTCTGCAAAAGATGTCCTCGACGAAAACGACAGAGTGCACACCCCTCCACG 50617
Dy	100	GTCCTCCCTCAACCAAGTGTCCCCCAAGAGCCAGACATGATCTGCTCCACGTTCTCTC 159
Db	50618	TTCCTCCCTCAACCAAGTGTCCCCCAAGAGCCAGACATGATCTGCTCCACGTTCTCTC 50677
Dy	160	TGCGTGTGCCCCCAAGCTCTGAGGCTCTGTGACCTTACTGCTCCAGAGGCGCTGCTCTGA 219
Db	50678	TGATGTGCTCTTAAGCTCCGAGGAGCTGTGAGCCCAAGTCTTGAAGTGTGCTGCTGAG 50737
Dy	220	CCACCAACGAGC-----GCCACACCGATGCGGAGCGCCAGAGGCCCACTCCTGTGACAG 273
Db	50738	CCACCAACGAGCATTGAGGTCCTATTAAGCGGCGCCAGAGATCCTGTTGACAG 50797
Dy	274	GCGCAGTGTGACGAAAGCGGAGCTCTGCTGCTGCTGCAACGTTCTGGGAGCTGTGCTG 333
Db	50798	GCGCAGTGTGACGAAAGCGGAGCTCTGCTGCTGCTGCAACGTTCTGGGAGCTGTGCTG 50857
Dy	334	ATCCAGATCCTGATGTCGTCGACGACAAAGCATCTTTGAGAGAAACAAGATCCCAAGAGCCA 393
Db	50858	ACCTGATCCTGATGTCGTCGACGACAAAGCATCTTTGAGAGAAACAAGATCCCAAGAGCCA 50917
Dy	394	AGAACACGCCAT-CTGACGCAATGCTTCCATATACCTCTTGTACTTTACAGAGCTG 452
Db	50918	GGAAGATCTCATCTGATGCAATGCTTTCCAGATATCCTGTTCTGGGTTTACAAAGCTG 50977
Dy	453	AGCTGAGAGTTTCTCTGTGGGGGATGTAGCTCTCCCAAGAGCATTTCTGTTTATG 512
Db	50978	AGCTGAGAGTTTCTCTGTGGGGGATGTAGCTCTCCCAAGAGCATTTCTGTTTATG 51037
Dy	513	TACAGATGCAATATCTCCCTTACCCCTGTACTCTGCAAGATTTGGCAGATGCTGTGCC 572
Db	51038	TATGATCTCACA-CTTCCCTCATCTCCCTGCCCCCTGCCAAGATTGCAATTCCTGTGC 51096
Dy	573	CAACTCTGTAAAAAAGATTAAGTCCGT 600
Db	51097	CTAACCTGTCAAAAAAATTAAGTCTTT 51124
RESULT 7		
LOCUS	AL139247	
DEFINITION	Human DNA sequence from clone RP1-52010 on chromosome 1, complete sequence.	
ACCESSION	AL139247	
VERSION	AL139247.12	GI:16444666
KEYWORDS	HTG.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS
TITLE
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 117084)
Thomas, D
Direct Submission
Submitted (24-Oct-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Oct 25, 2001 this sequence version replaced gi:15021286.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WormPep; Information on the WormPep
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RPI-52J10 is from the library RPI-1 constructed by the group of
Pleier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2

COMMENT

This sequence is the entire insert of clone RPI-52J10 The true left
end of clone RPI-43017 is at 61046 in this sequence. The true right
end of clone RPI-11013 is at 14156 in this sequence.
Location/Qualifiers
1..117084
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="RPI-52J10"
/clone_id="RPI-1"
46310..46544
/note="Sequence from uni-directional dGTP big dye
terminator reads only."
69467..69488
/note="Sequence from uni-directional dGTP big dye
terminator reads only."
misc_feature
35165 a 24224 c 24423 g 33272 t
BASE COUNT
ORIGIN

FEATURES
source
1..117084
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="RPI-52J10"
/clone_id="RPI-1"
46310..46544
/note="Sequence from uni-directional dGTP big dye
terminator reads only."
69467..69488
/note="Sequence from uni-directional dGTP big dye
terminator reads only."
misc_feature
35165 a 24224 c 24423 g 33272 t
BASE COUNT
ORIGIN

Query Match 59.8%; Score 360.4; DB 9; Length 117084;
Best Local Similarity 80.8%; Pred. No. 4e-76;
Matches 459; Conservative 0; Mismatches 101; Indels 8; Gaps 3;
40 CTCACCTCTGCGAAGATGCTCTGCGACGAGACGACGAGTCCACCCCAACCA 99
203 CTGACATCTCTCAAGAGTCTCTGCGACGAGACGAGTCCACCCCAACCA 262
100 GTGTCCTCTCAAGAGTCTCTGCGACGAGACGAGTCTGCTGCTCAAGTCTCTC 159
263 TTGTCCTCTCAAGAGTCTCTGCGACGAGACGAGTCTGCTGCTCAAGTCTCTC 322
160 TGGCTGTCCTCAAGAGTCTCTGCGACGAGACGAGTCTGCTGCTCAAGTCTCTC 219
323 TGACTGTGCTCTCAAGTCTCTGCGACGAGACGAGTCTGCTGCTCAAGTCTCTC 382
220 CCACACAGGC-----GCACACACGAGTCTGCGACGAGACGAGTCTGCTGCTCAAG 273

Db 383 CCACACAGGCACTTCAAGTCCCATCATGATCCGCGCCGACAGATCCACTCTGTGACNG 442
Qy 274 GGGCAATGTGTCAAGAGCGGGGGCTCTGCTGTGCTGCAAGTCTGCGGGCTGTGCTG 333
Db 443 GGGCAATGTGTCAAGAGCGGGGGCTCTGCTGTGCTGCAAGTCTGCGGGCTGTGCTG 502
Qy 334 ATCCATCTCTATGATGAGCAAGAGTCTTGGAGGAAACAAGATCCCAAGAGGCCA 393
Db 503 ACTGATCTCTATGATGAGCAAGAGTCTTGGAGGAAACAAGATCCCAAGAGGCCA 562
Qy 394 AGAACAGCCCAT-CTGACGATGCTTCCCAATPACCTCTTGTGATTTCAAGGCTG 452
Db 563 GGAAGAGCTTCAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATG 622
Qy 453 AGCTGAGGTTTCCGTGCGGGGATCTGAGTCTCTCCCAAGGACATCTTGTGTTTNG 512
Db 623 AGCTGAGGTTTCCGTGCGGGGATCTGAGTCTCTCCCAAGGACATCTTGTGTTTNG 682
Qy 513 TACAGATGTCAATATGCTCCCTTACCTGACCTGCAAGATGAGTGTGCTGTC 572
Db 683 TACAGATGTCAAT-CTTCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 741
Qy 573 CAACCTCTTAAAGATTAAGTTCCT 600
Db 742 CTAACTCTGCAAAATTAAGTTCCT 769

RESULT 8
AC140641/C
DEFINITION
Macaca mulatta clone CH250-288A21. *** SEQUENCING IN PROGRESS ***
AC140641
VERSION
AC140641.5 GI:29650194
KEYWORDS
HTG; HTGS PHASE2; HTGS PGI.
SOURCE
Macaca mulatta (rhesus monkey)
ORGANISM
Macaca mulatta

REFERENCE
AUTHORS
TITLE
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
Cercopitheidae; Macaca.
1 (bases 1 to 228396)
Cauro, M. and Milosavljevic, A.
Pooled genomic indexing (PGI): mathematical analysis and experiment
design
(in) Guigo, R. and Gelfand, D. (Eds.);
ALGORITHMS IN BIOINFORMATICS, SECOND INTERNATIONAL WORKSHOP, WABI
2002, ROME, ITALY, SEPTEMBER 17-21, 2002, PROCEEDINGS, 10-28;
Springer (2002)
2 (bases 1 to 228396)

REFERENCE
AUTHORS

Milosavljevic, A., Sodergren, E., Cauro, M., Li, B., Jackson, A.R.,
Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Albrooks, S.L.,
Amaratunga, H.C., Are, J.R., Ayala, M., Banks, T., Barbata, J.,
Benton, J., Blum, K., Blankenburg, K., Bonin, D., Bouck, U.,
Bowle, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chin, D., Chowdhury, I., Christopoulos, C., Cleveland, C.D.,
Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy, Carroll, J., Dedert, J.A., Delaney, K.R., Delgado, O.,
Dem, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Egan, A., Barnhart, C., Edwards, C.C.,
Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D.,
Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J.,
Garcia, A., Garner, T., Garza, N., Gill, R., Gottell, J.H., Guevara, M.,
Gunnarsson, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K.,
Hart, M., Havlik, P., Hawes, A., Hernandez, J., Hernandez, O.,
Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Homisi, F.,
Howard, S., Huber, J., Hulik, S., Hume, J., Ischikawa, I., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Joiliver, S., Joudah, S.,
Karlsso, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kuresh, A., Landry, N., Leal, B., Lee, E., Lewis, J.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W.,
Luissege, H., Lorzado, R., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Matonde, I., Martin, R.,

```

Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P.,
Meador,M., Mei,G., Merscher,S., Metzger,M., Miller,A., Miner,G.,
Miner,Z., Mitchell,T., Monabdat,K., Montgomery,K.T., Morgan,M.,
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S.,
Ogih,M., Okunomiya,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Plikus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,
Ruitz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
Shooshita,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A.,
Stanley,H., Stone,H., Sutton,A., Swalek,A., Tabor,P., Tamerila,A.,
Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Wallington,S., Williams,G., Williamson,A.,
Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.F., Zhou,J.,
Zorrilla,S., Kuchelapatti,R., Weinstein,G. and Gibbs,R.
Direct Submission
Unpublished
3 (bases 1 to 228396)
Worley,K.C.
Direct Submission
Submitted (26-FEB-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 228396)
Worley,K.C.
Direct Submission
Submitted (09-APR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 9, 2003 this sequence version replaced gi:29568000.

----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: LAAY
Center clone name: CH250-288A21
----- Summary Statistics
Sequencing vector: plasmid
Chemistry: Dye-terminator Big Dye, 100% of reads
Consensus quality: 9951 bases at least Q40
Consensus quality: 12209 bases at least Q30
Consensus quality: 14605 bases at least Q20
Estimated insert size: 99169; sum-of-coverage estimation
Quality coverage: 1x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: The contigs are based on the application
* of the Pgi method using the Human genome (NCBI build 31)
* as the comparative genome.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 228396: contig of 228396 bp in length.
Location/Qualifiers
1..228396
/organism="Macaca mulatta"
/mol_type="genomic DNA"
/db_xref="taxon:9544"
/clone="CH250-288A21"
1..228396
/notes="assembly_name:CH250-288A21.1A
CONFIDENCE:_0.83"

FEATURES
source
misc_feature

TITLE
REFERENCE
AUTHORS
JOURNAL
COMMENT
```

```

BASE COUNT      5443 a      4300 c      4314 g      5247 t      209092 others
ORIGIN
Query Match
Best Local Similarity 88.0%; Pred. No. 3,76-37;
Matches 227; Conservative 0; Mismatches 30; Indels 1; Gaps 1;

QY 341 TCTGATCTGAGACACACGATCTTTGGAGAAAACAGATCCCAAGGCCCAAGACAG 400
DB 153918 TCTTATATCGAAGATTAAGCGATCTTTGGAGAAAACAGATCCCAAGGCCCAAGACAG 153859
QY 401 CCCCAT-CTGACGATGCTCCCATATACCTCTTGATTTACAGGCTGAGCTGGA 459
DB 153858 CCCCATCTGACGATGCTCCCATATACCTCTTGATTTACAGGCTGAGCTGAG 153799
QY 460 GATTTCCTGTGGGGAGATCTAGCTCTCCCAAGAGGACATTTGTTTATGTACAGGA 519
DB 153798 GATTTCCTGTGGGGAGATCTAGCTCTCCCAAGAGGACATTTGTTTATGTACAGGA 153739
QY 520 TGTATATGTCCTCCCTACCTCTGATCTGACCTGACAGATTTGGAGGCTTTGACCACTC 579
DB 153738 TGTATATGTCCTCTTCCCTGTACCTGCCAAGATTTGGAGGCTTTGACCACTC 153679
QY 580 GTAAAAAAGATTAAGTTC 597
DB 153678 GTCAAAAATAAAGTTTC 153661

RESULT 9
AC132274/c 199220 bp DNA linear HTG 03-SEP-2002
AC132274/c Mus musculus chromosome UNK clone RP24-341121, WORKING DRAFT
DEFINITION
AC132274.1 GI:22657732
VERSION
AC132274.1 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
McPherson,J.D. and Waterston,R.H.
TITLE
The sequence of Mus musculus clone
JOURNAL
Unpublished
2 (bases 1 to 199220)
AUTHORS
McPherson,J.D. and Waterston,R.H.
REFERENCE
Submitted (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL
Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: W.BB0341121
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator Big Dye, 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 19711 bases at least Q40
Consensus quality: 197384 bases at least Q30
Consensus quality: 197515 bases at least Q20
Insert size: 18300; agarose-fp
Insert size: 198257; sum-of-coverage
Quality coverage: 17.62 in Q20 bases; agarose-fp
Quality coverage: 13.20 in Q20 bases; sum-of-coverage
* NOTE: This is a 'working draft' sequence. It currently
```


DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
Rattus norvegicus clone CH230-unknown, *** SEQUENCING IN PROGRESS	AC137251	GI:25073159	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.	Rattus norvegicus (Norway rat)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
*** 3 unordered pieces.					
REFERENCE					
AUTHORS					
Munry,D,Marle, Metzker,M,Lee, Abramzon,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D, Arayaibeche,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Behnmed,F, Biwalto,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Chen,A, Chacko,V, Chavez,D, Chen,G, Chen,R, Chen,Y, Chter,A, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davis,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Hayes,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gablis,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Georgeogis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Gunaratne,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huijck,S, Hume,J, Idelbird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levay,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,M, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenshweta,L, Loulsged,H, Lozdo,R,U, Lu,X, Ma,J, Meheswari,M, Mahindaratne,M, Mahmoud,K, Mallory,K, Mangum,A, Mangun,B, Mapa,P, Martin,K, Martin,R, Martinez,E, Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L, Nervens,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nsakoeleneh,O, Okunou,G, Olarnunagoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perre,A, Perez,L, Pfannkuch,C, Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L, L, Pnazo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savery,G, Scheerer,S, Scott,G, Shatman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sison,I, Sitter,C,D, Smitz,D, Sneed,A, Sodergren,E, Song,X,-Z, Sorelle,R, Sosa,J, Steinle,M, Strong,R, Sutton,A, Svatek,A, Taboc,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willison,R, Wlaczek,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinsteck,G, and Gibbs,R,A.					
REFERENCE					
AUTHORS					
Unpublished					
2 (bases 1 to 195503)					
Rat Genome Sequencing Consortium.					
Direct Submission					
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,					
COMMENT					

Individual sequence contigs are ordered and oriented, and separately sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KZ12
Center clone name: CH230-unknown

```
----- Summary Statistics -----
Assembly program: Phrap; version 0.90329
Consensus quality: 164089 bases at least Q40
Consensus quality: 166085 bases at least Q30
Consensus quality: 167609 bases at least Q20
Estimated insert size: 171122; sum-of-configs estimation
Quality coverage: 6x in Q20 bases; sum-of-configs estimation
```

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.igsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

182074	180273: contig of 180273 bp in length
180374	180373: gap of unknown length
181900	181899: contig of 1526 bp in length
181999	181999: gap of unknown length
182000	195503: contig of 13504 bp in length.

```

LOCUS       50392 a 34104 c 34657 g 48942 t 27408 others
SOURCE      Rattus norvegicus
  ORGANISM  Rattus norvegicus
            Mol. type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-unknown"
            1..1873
            /note="wgs contig"
            174826..176310
            /note="wgs contig"

```

Length	195503;
Score	185.4; DB 2;
Similarity	77.0%; Pred. No. 3.9e-34;
Mismatches	66;
Indels	19;
Gaps	4;

6 TCCTGCAGATGTCCTGCAGCAGAACGACAGCAGTGCCAACCCCAACCAAGTGTCC 105

CTCA-----CCCAAGTGTCCCCCAAAGGCCAGTACAGTGTCTGCTCCAGC 153
9 TCCAGCAGATGTCCTGCCAGCAGAGCCAGAAAGCAGTGCACGCTCTCCCAAGTGCCC 1597

9 CTCCCCGAAGTCTCCCCAAGTCCCCCAAGGACAGCACAGTGTCTGCGCTGCAGC 1603

9 CTCTTCTGCTGTGCTACAGCTCTGGGGGCTGCAAGTGTCCCCAGCTCTGAGGGAGGCTG 1609

1 CTTCTGAACCAACAGGCGC---CACCACCATGCCGGCCAGAAGCCAACTCCTG 267
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9 CTGCCTGACCACCAAGGCGCAGTCCCAAGATGACGGCGAGAGCTTCAGTTCGTG 1615

8 TGACAGGGGCAAGTGGTCAGCAAGGCGGGGCTCTGCTGCTGCACCGGTTCTGGGGGCTG 327

Db 16159 TGACCGGCGAGCTGCTACGACGCTGGGGGCTCAGGCTGTGTGTACACCTCTGGGGGCTG 16218

Qy 328 CTGCTGATCCAGATCCGATGCTGAGACAAAGCGATCTTTGGAGAAAACAATCCCAAG 387

Db 16219 CTGCTAATCTGGAACATGATCTGTGATGACAGGA-GTCGGGAGAGAAAAAGATCCCAAA 16277

Qy 368 AGGCCAAGA 396

Db 16278 GGTCAAGA 16286

RESULT 12

AC095371

LOCUS

DEFINITION Rattus norvegicus clone CR230-1G20, *** SEQUENCING IN PROGRESS ***

AC095371

VERSION AC095371.4 GI:24941586

KEYWORDS HTG, HTGS, PHASE1, HTGS DRAFT, HTGS_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 261316)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Albrooks,S.L., Amarunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonini,D., Bouck,J., Bowe,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Burch,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davis,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Derr,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Hombl,F., Howard,S., Huber,J., Huijy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,T., Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,T., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulsged,H., Locado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheswari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Maesey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwuo,S., Ogih,M., Okumura,G., Ogrunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Prime,E., P.L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Severy,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Soederren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Swalek,A., Tabor,P., Tamerisa,X., Tang,H., Tasey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Uman,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,C., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleceyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zortilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

Unpublished

Direct Submission

2 (bases 1 to 261316)

Worley,K.C.

REFERENCE

JOURNAL

AUTHORS

TITLE

JOURNAL

Submitted (16-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

3 (bases 1 to 261316)

AUTHORS

Worley,K.C.

TITLE

Direct Submission

Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 13, 2002 this sequence version replaced gi:23269262. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: TUIT

Center clone name: CR230-1G20

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 228141 bases at least Q40

Consensus quality: 225987 bases at least Q30

Consensus quality: 231269 bases at least Q20

Estimated insert size: 24114; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

----- NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES	source
1	2233: contig of 3233 bp in length
3234	3333: gap of unknown length
3334	218784: contig of 215451 bp in length
218785	222557: gap of unknown length
218885	222557: contig of 3373 bp in length
222558	223507: gap of unknown length
222559	223507: contig of 1150 bp in length
223508	223607: gap of unknown length
223509	223607: contig of 1000 bp in length
224608	224607: gap of unknown length
224708	261316: contig of 36603 bp in length.

Location/Qualifiers

1. 261316

/organism="Rattus norvegicus"

/mol_type="Genomic DNA"

/db_xref="taxon:10116"

/clone="CR230-1G20"

1. 1274

/note="wgs_contig"

3334. 4756

/note="wgs_contig"

218885. 219956

/note="wgs_contig"

220761. 222257

/note="wgs_contig"

BASE COUNT 67615 a 48827 c 47477 g 67871 t 29526 others

ORIGIN

[illegible]

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinley, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Muridasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Naokoletneh, O., Okunogun, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Pioger, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.,
Pizzo, M., Quirio, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, K., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shasman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaj, D.,
Sneed, A., Sodeghren, E., Song, X.-Z., Sorrell, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Szatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tilgney, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseña, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, Y., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausen, A., Weiser, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Wernscock, G., and Gibbs, R.A.
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 347801)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23269758.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYDP
Center clone name: CH230-364N10
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 314830 bases at least Q40
Consensus quality: 318638 bases at least Q30
Consensus quality: 321617 bases at least Q20
Estimated insert size: 328283; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

